Request to Archive
With The National Centers for Environmental Information
For Global Precipitation Climatology Project (GPCP)
Provided by NASA

2014-11-12

This information will be used by NCEI to conduct an appraisal and make a decision on the request.

1. Who is the primary point of contact for this request?

Howard Diamond
NCDC
Director, World Data Center for Meteorology
+1-301-427-2475
howard.diamond@noaa.gov

2. Name the organization or group responsible for creating the dataset.

NASA/GSFC Mesoscale Atmospheric Processes Laboratory; David Bolvin, GPCP Dataset Curator <david.t.bolvin@nasa.gov>; Tel: (301) 614-6323

3. Provide an overview summarizing the scope of data you want to archive. Describe the outputs, data variables, including their measurement resolution and coverage.

The Global Precipitation Climatology Project (GPCP) was established by the World Climate Research Program (WCRP) and subsequently attached to the Global Energy and Water Exchange program (GEWEX) to address the problem of quantifying the distribution of precipitation around the globe over many years. The general approach is to combine the precipitation information available from each of several sources into a final merged product, taking advantage of the strengths of each data type. The passive microwave estimates are based on Special Sensor Microwave/Imager (SSMI) and Special Sensor Microwave Imager/Sounder (SSMIS) data from the series of Defense Meteorological Satellite Program (DMSP, United States) satellites that fly in sun-synchronous low-earth orbits at 6 a.m./p.m. The infrared (IR) precipitation estimates are computed primarily from geostationary satellites (United States, Europe, Japan), and secondarily from NOAA-series polar-orbiting satellites (United States). Additional low-Earth orbit estimates include Atmospheric Infrared Sounder (AIRS data from the NASA Aqua, and Television Infrared Observation Satellite Program (TIROS) Operational Vertical Sounder (TOVS) and Outgoing Longwave Radiation (OLR) Precipitation Index (OPI) data from the NOAA series satellites. The precipitation gauge data are assembled and analyzed by the Global Precipitation Climatology Centre (GPCC) of the Deutscher Wetterdienst.

The GPCP has promoted the development of an analysis procedure for blending the various estimates together to produce the necessary global gridded precipitation fields. The currently operational procedure is described in Adler et al. (2003) and Huffman et al. (2009), and has been used to produce the GPCP Version 2.2 Combined Precipitation Data Set, covering the period January 1979 through the present (with some delay). The primary product in the Version 2.2 dataset is a combined observation-only dataset, that is, a gridded analysis based on gauge measurements and satellite estimates of precipitation. There are a total of 27 fields in the data set providing information from the individual and intermediate estimates, including estimates of RMS random error.

The data set archive consists of yearly unformatted REAL*4 binary files with ASCII headers, each of which holds 12 monthly fields. Each file occupies almost 0.5 MB. The grid on which each field of values is presented is a 2.5°x2.5°

latitude--longitude (Cylindrical Equal Distance) global array of points. It is size 144x72, with X (longitude) incrementing most rapidly West to East from the Prime Meridian, and then Y (latitude) incrementing North to South. Grid edges are placed on whole- and half-degree values:

First point center = (88.75°N,1.25°E)
Second point center = (88.75°N,3.75°E)
Last point center = (88.75°S,1.25°W)

Missing values are denoted by the value -99999., and the units on the fields depend on the variable. For example, rainfall is carried as mm/day.

4. What is the time period covered by the dataset? (YYYY-MM-DD, YYYY-MM or YYYY)
From 1979-01-01
Ongoing as continuous updates to the data record

5. Edition or version number(s) of the dataset:
Version 2.2 for GPCP monthly data; and Version 1.2 for GPCP One-Degree Daily (1DD) data.

6. Approximate date when the dataset was or will be released to the public:
1990-01-01

7. Who are the expected users of the archived data? How will the archived data be used?
The users of the GPCP data constitute a wide array of global scientists involved in climatic precipitation studies. In addition to being used by numerous investigators, the data also constitute an input to NCDC's Climate Data Records project.

8. Has the dataset undergone user evaluation and/or an independent review process? Did NCEI participate in design reviews?
This data has undergone peer review by the scientific community, and the references are as follows:

Adler, R.F., G.J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, E. Nelkin 2003: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). J. Hydrometeor., 4,1147-1167.

Other citations are available at: http://www1.ncdc.noaa.gov/pub/data/gpcp/gpcp-v2.2/doc/gpcp_citation_list.pdf

Huffman, G.J, R.F. Adler, D.T. Bolvin, G. Gu 2009: Improving the Global Precipitation Record: GPCP Version 2.1. Geophys. Res. Lett., 36,L17808, doi:10.1029/2009GL040000.

9. Describe the dataset's relationship to other archived datasets, such as earlier versions or related source data. If this is a new version, how does it improve upon the previous version(s)?
N/A

10. List the input datasets and ancillary information used to produce the data.
The Global Precipitation Climatology Project (GPCP) was established by the World Climate Research Program (WCRP) and subsequently attached to the Global Energy and Water Exchange program (GEWEX) to address the problem of quantifying the distribution of precipitation around the globe over many years. The general approach is to combine the precipitation information available from each of several sources into a final merged product, taking

advantage of the strengths of each data type. The passive microwave estimates are based on Special Sensor Microwave/Imager (SSMI) and Special Sensor Microwave Imager/Sounder (SSMIS) data from the series of Defense Meteorological Satellite Program (DMSP, United States) satellites that fly in sun-synchronous low-earth orbits at 6 a.m./p.m. The infrared (IR) precipitation estimates are computed primarily from geostationary satellites (United States, Europe, Japan), and secondarily from NOAA-series polar-orbiting satellites (United States). Additional low-Earth orbit estimates include Atmospheric Infrared Sounder (AIRS data from the NASA Aqua, and Television Infrared Observation Satellite Program (TIROS) Operational Vertical Sounder (TOVS) and Outgoing Longwave Radiation (OLR) Precipitation Index (OPI) data from the NOAA series satellites. The precipitation gauge data are assembled and analyzed by the Global Precipitation Climatology Centre (GPCC) of the Deutscher Wetterdienst.
See also:
http://mirador.gsfc.nasa.gov/cgi-bin/mirador/presentNavigation.pl?tree=project&project=TRMM&dataGroup=Gridded
http://precip.gsfc.nasa.gov/trmm_comb.html
http://precip.gsfc.nasa.gov/gpcp_daily_comb.html

11. List web pages and other links that provide information on the data.
http://precip.gsfc.nasa.gov/

12. List the kinds of documents, metadata and code that are available for archiving. For example, data format specifications, user guides, algorithm documentation, metadata compliant with a standard such as ISO 19115, source code, platform/instrument metadata, data/process flow diagrams, etc.
1. http://www1.ncdc.noaa.gov/pub/data/gpcp/1dd-v1.2/1DD_v1.2_doc.pdf
   http://www1.ncdc.noaa.gov/pub/data/gpcp/gpcp-v2.2/doc/V2.2_doc.pdf
   http://www1.ncdc.noaa.gov/pub/data/gpcp/pentad/documentation/gpcp_pen_v2.2.doc

13. Indicate the data file format(s).
1. Unformatted binary files with ASCII header

14. Are the data files compressed?
No

15. Provide details on how the files are named and how they are organized (e.g., file_name_pattern_YYYYMM.tar in monthly aggregations).
(1) Pentad Data: gpcp_pen_v2.2.lnx.YYYY
(2) Intermediate files per http://www1.ncdc.noaa.gov/pub/data/gpcp/gpcp-v2.2/doc/V2.2_doc.pdf:
gpcp_v2.2_nnn.YYYY, where nnn = eag, ega, ems, eop, est, nga, ngp, nsc, nse, nss, pag, pga, pgp, pmp, pop, psc, pse, pss, pst, pty, ssc, and sst
(3) Daily 1DD data: gpcp_1dd_v1.2_p1d.YYYYMM
(4) Satellite Gauge Precipitation: gpcp_v2.2_psg.YYYY
(5) Satellite Gauge Error: gpcp_v2.2_esg.YYYY

16. Explain how to access sample data files and/or a file listing for previewing. If it is not available now, when will it be available?
Previews of the data are posted on the NASA site at http://precip.gsfc.nasa.gov/

17. What is the total data volume to be submitted?
Historic Data: all historic data or data submitted as a completed collection.
Total Data Volume: 2.2GB

Number of Data Files: 977

Continuous Data: data volume rate for a continuous data production.

Total Data Volume Rate: 61MB per Month
Data File Frequency: 1 per Month
Data Production Start: 2015-01-05

18. Are later updates, revisions or replacement files anticipated? If so, explain the conditions for submitting these additional data to the archive.

Updated data are made available on a monthly basis.

19. Describe the server that will connect to the ingest server at NCEI for submitting the data.

Physical Location: At NASA's Goddard Space Flight Center
System Name: N/A
System Owner: N/A
Additional Information: N/A

20. What are the possible methods for submitting the data to NCEI? Select all that apply.

1. FTP PULL

Data made avail for pull at http://precip.gsfc.nasa.gov per an e-mail message from dataset curator

21. Identify how you would like NCEI to distribute the data. Web access support depends on the resources available for the dataset.

22. Will there be any distribution, usage, or other restrictions that apply to the data in the archive?

No known constraints apply to the data.

23. Discuss the rationale for archiving the dataset and the anticipated benefits. Mention any risks associated with not archiving the dataset at NCEI.

The GPCP dataset is a foundational one involving global precipitation data that are used in a number of research efforts. There has been a historical relationship between NCDC's staging of the data with NASA that dates back to the very first release of the GPCP data sets in the late 1990's, The current version of the GPCP dataset are used as part of the in the NOAA Climate Data Records program.

24. Are the data archived at another facility or are there plans to do so? Please explain.

No

25. Is there an existing agreement or requirement driving this request to archive? Have you already contacted someone at NCEI?

The official archive for GPCP data has been the World Data Center for Meteorology (WDC), Asheville dating back to the 1990s, and the WDC has been maintained at NCDC during that entire time. NASA's technical documentation traces back to the template that NCDC required NASA to use in that first posting, working with the NCDC contact at the time, Al McNab. The original agreements were probably more informal in nature, but after over 20 years of performing this service and via two different WDC Directors (Gus Shumbera and Howard Diamond), this originally informal agreement has done very well. The intent here is to formalize this long-standing data service provided to the global climate community

26. Do you have a data management plan for your data?
No

27. Have funds been allocated to archive the data at NCEI?
No

28. Identify the affiliated research project, its sponsor, and any project/grant ID as applicable.
N/A

29. Is there a desired deadline for NCEI to archive and provide access to the data?

Archive by:	2015-01-05
Accessible by:	2015-01-05

30. Add any other pertinent information for this request.
NCDC's relationship to posting GPCP data goes back to the early 1990s when agreements for archiving data were more informal. Given that the global climate community has come to depend on this data, and given its importance, it is hoped that the formal archiving of this data can be approved as a grandfathered dataset that the World Data Center for Meteorology has been looked to for hosting this dataset. While I have noted an archive and accessible data of 2015-01-05 in question 28, in reality this data continues to be made accessible, and this request is intended to formalize the archiving of this data. Therefore, the date for this formal archiving can be done later than this date.